

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

In Chapters 1 through 5, we described the main activities conducted in the first year of the HSEE evaluation and summarized the results of these activities to date. In this chapter, we present the main conclusions that we draw from these results and list recommendations for further HSEE development based on these conclusions.

General Conclusions

Conclusion 1. A great deal has been accomplished since the HSEE development was launched.

As specified in the enabling legislation, an *HSEE Standards Panel* was appointed by the Superintendent with review and approval by the Board. The Panel has met monthly and has reviewed and discussed a wide range of issues from test specifications to administration (particularly required accommodations) and reporting issues. Along with a technical committee of content and measurement experts, the Panel recommended test specifications incorporating additional information on test blueprints proposed by the development contractor.

Contracts were let for both the development and the evaluation of the HSEE. This was not a small accomplishment, as contractors were initially hesitant to bid on the development effort due to its ambitious demands and tight timelines. In the interim, Department staff worked hard to support the HSEE Panel and identify appropriate items to keep the project on schedule. Since the contracts were let in January and February of this year, Department staff have provided oversight and guidance to each effort and have continued to be heavily engaged in coordinating the efforts of the contractors, the Panel, and others.

A large number of *test items have been developed* and shepherded through an extensive review process. The process included one or more rounds of review for technical, editorial, bias and sensitivity, and policy issues, including a review by the HSEE Panel. A total of 362 English Language Arts items and 396 Mathematics items survived this process and were included in field test booklets.

The field test was conducted as scheduled. The development contractor identified a broadly representative set of schools, recruited participation, and ended up testing more than 7,500 10th grade students from these schools. Each student completed a lengthy test booklet containing 99 or 102 test items and nearly all of the students answered all or nearly all of the items.

Conclusion 2. Development efforts to date have been highly successful.

Agreement was reached by the Panel, the developers, and the department concerning the detailed *contents of the exam*. The contents that will be proposed for Board adoption reflect a hard-won consensus among the different perspectives and positions reflected on the Panel and across other stakeholder groups.

The quantity and, most importantly, quality of the items developed to date are high. The HSEE Panel and its Technical Committee reviewed these items and agreed that they were each acceptable measures of the targeted standards. We conducted an independent review of all of the items by panels of teachers and curriculum specialists and, with few exceptions, they agreed that the items were reasonably aligned to the proposed test standards.

Relatively few items were flagged for statistical problems in the field test. A high proportion of the items were within an acceptable range of difficulty, differentiated between high and low performing students as measured by the other items, and did not function differently for minority students. The positive results from the field test validate the development and review process conducted by the contractor and the Department.

Preliminary evidence suggests that scores based on the items developed to date would probably lead to accurate pass/fail decisions across a range of plausible minimum scores. (See the analyses reported in Chapter 4.) Ultimately, however, decisions about the passing scores and the required level of accuracy in classifying students with respect to these scores require policy judgments.

Conclusion 3. Much more needs to be done before operational administration can begin.

Several important questions about the exam and its use remain to be answered. Some of the possible answers would appear to make operational administration in winter and spring 2001 infeasible. Some of the more critical questions are:

1. *Will the test blueprints, specifying the number of items to be included for each of the targeted standards, be approved as currently recommended?* If significant changes are made, it may be too late to develop and field test items in time for inclusion in the 2001 test forms.
2. *How many different test forms will be needed for the 2001 administrations?* Separate March and May administrations are planned, in order to accommodate school schedules and year-round schools, in particular. Thus a minimum of two test forms will be needed, and one or more backup forms may also be needed in case of security breaches. Based on the items developed to date, it may be possible to develop two independent forms, but if more forms are needed, it is not clear that current deadlines can be met.
3. *How will the minimum passing score be determined?* It seems unlikely that this question can be answered until one or more operational test forms have been assembled. The procedures used with the National Assessment of Educational Progress and many state assessments where standards-based reporting is employed require some normative information. At some point in the process prior to final approval, information on consequences, meaning the proportion of students who will pass, is examined. Unless reporting of the 2001 results is significantly delayed (and the legislation requires reporting within 8 weeks), the passing standards will have to be set before the operational test form(s) are administered to any students, let alone a

representative (norming) sample. It may be possible to “simulate” passing rates based on individual item data from the field tests. Such simulated passing rates require a number of statistical assumptions (more complex than the assumptions underlying sampling with the U.S. Census, which were rejected by Congress). There are no suggestions on the table for collecting normative data for use in setting passing scores and still meeting the schedule for the 2001 administrations. We have no suggestions either.

4. *What information, if any, will be reported to students, parents, and teachers in addition to the overall pass or fail outcome for English/Language Arts and for Mathematics?* Time will be required to develop and test procedures for printing score reports, so final decisions about the content of these reports will likely be needed two or more months before the first administration.
5. *How will scores from different test forms be put on a comparable scale?* A minimum of two forms will be needed (i.e., one for March 2001 and another for May 2001, and possibly backup forms will be required in case of security breaches. Due to inevitable minor differences in item difficulty, it is not reasonable to assume that a given number correct score represents the same level of proficiency for different test forms. Equating based on common items and/or administration of the different test forms to strictly comparable samples of students is required to equate the scores from these forms. Equating plans must be developed, reviewed, and approved before additional test forms can be assembled, meaning well in advance of operational use of these forms.
6. *What specific accommodations will be provided on each of the two tests for special needs students and for English-language learners?* Final approval of proposed accommodations (as specified in EC 60850g) is needed along with evidence that scores for students tested with these accommodations are comparable to scores for students with similar proficiencies who do not require such accommodations.

Even if the answers to these questions do not render current plans for the first operational administration infeasible, the schedule is daunting. A second field test scheduled for this fall is needed to fill in some gaps in coverage of the test standards and to support the construction of additional test forms. Given the need to complete the fall field test, it seems unlikely that the first test forms can be assembled before December, and final copy for these forms will probably need to go to the printers in January. This leaves little time for review or any other possible delays.

Conclusion 4. Available evidence suggests reason to be concerned about how well students are currently prepared to meet the standards of the assessment.

Feedback from principals and teachers, ratings by our panels of educators, and the percent of students who answered each item correctly in the field test all suggest that half or more of today's 10th graders would fail one or both of the HSEE tests, particularly the mathematics test. On the one hand, this evidence underscores the need for the HSEE or some other interventions to raise levels of achievement relative to the targeted standards. On the

other hand, if a significant number of high school students are denied a diploma on the basis of this exam, there are likely to be many negative, as well as positive consequences.

Results from the field test were consistent with 1999 STAR results and results from other examination in suggesting that minority students, special needs students, and English-language learners are particularly at risk of failing the HSEE.

Recommendations

In addition to reporting on consequences of the new HSEE requirements, the Education Code (60855c) requires the evaluators to “include recommendations to improve the quality, fairness, validity, and reliability of the examination.” Based on the results of the evaluation to date, as summarized in the general conclusions above and the more detailed findings in the body of this report, we offer one general and three more specific recommendations at this time. In making these recommendations, we are mindful of the need to maintain the independence of this evaluation. To the extent that evaluators become proponents for specific plans, they lose at least the appearance of independence when evaluating these same plans. Consequently, we will avoid proposing our own plans for processes where the Department, the Panel, and the development contractor have not yet finalized their own plans (e.g., setting minimum passing scores). We shall have more to say about those processes in subsequent reports, after initial plans are proposed by others.

General Recommendation

The State Board of Education, Legislature, and Governor should give serious consideration to postponing full implementation of the HSEE requirement by 1 or 2 years.

There are two reasons why it might be important to delay full implementation of the exit examination, as was done recently in Maryland. First, as suggested by the list of decisions yet to be made in Conclusion 3 above, it may take more time to develop an assessment that fully meets professional and legal standards. For example, the timeline for reaching a decision on the minimum score needed to pass the test seems particularly problematic. A hasty or ill-informed decision about minimum passing scores will be hard to correct once students take the exam in 2001. The need for careful consideration and analyses of accommodations offered to special needs students is another point at which additional time prior to full operational implementation may be needed. The current schedule offers little time to develop the accommodations and no time for additional research on the comparability of scores across different possible testing accommodations. We note that established testing programs with which we are familiar (e.g., the Medical College Admissions Test and the Armed Services Vocational Aptitude Battery) spend three to four years on the development of each new test form. To complete the development of operational HSEE test forms in 18 months, with most of the first 6 months spent in developing test specifications (which are even yet not final) is ambitious.

The second, and we think more important, reason for considering a delay is that schools will need more time to prepare students to meet the standards assessed by the HSEE. As

described in Chapter 2, the key legal issue in prior challenges to high stakes test is whether students have been provided adequate instruction in the material covered by the test. Current plans call for students and schools to be fully notified about the exam and its requirement this fall, as the first affected class (the Class of 2004) enters 9th grade. This will be too late to allow very significant changes in the 9th grade curriculum for these students. In addition, the 7th and 8th grade curriculum must be aligned to prepare students for new curriculum changes at the high school level and because some of the standards being assessed are taken from the 8th grade standards. The need for more time to prepare seems particularly acute for special needs students, as updating IEPs to cover the contents of the exam will take some time.

Evidence that a significant proportion of today's students would not pass the test summarized under Conclusion 4, suggests that schools may need more time to provide adequate preparation for all students and to develop and implement remediation programs for students who do not initially pass.

With a delay of at least a year, a fully developed form could still be administered to 9th and 10th grade students in a representative sample of schools in 2001. The result of this would be important information for setting passing rates and, after the rates are set, important feedback to participating schools about how well their students are prepared for the exam¹¹. This would provide schools both information and time to implement further improvements to the curriculum before holding the students, themselves, fully responsible.

More Specific Recommendations

The process of developing and reviewing items appears to be working well and we have no specific recommendations for improving these processes at this time. Plans for assembling and equating different test forms and for setting minimum passing scores are still being developed and so we will reserve comment until a later report. At this time, we offer specific recommendations in three areas: (1) clarifying the relationship of the HSEE to other statewide testing programs, (2) communicating with district and school personnel about the HSEE, and (3) gathering information specific to the HSEE on the appropriateness and effectiveness of different testing accommodations.

Specific Recommendation 1. The Department and the Board need to work together to clarify the relationships and differences among the different high school testing programs, most notably the HSEE, the standards-based STAR assessment, and the Golden State Examinations.

Responses to our surveys of principals and teachers indicate that the fact that the HSEE and the standards-based STAR assessment are both designed to test student achievement relative to the same content standards is confusing, at best. Concerns about coordination of these different testing programs were also expressed at the June meeting of the State Board. One important difference is that the HSEE is targeted at levels of proficiency considered to be the minimum for high school graduation while the standards-based STAR assessment

¹¹ Schools will get important feedback about the performance of their students relative to the broader set of state curriculum standards from the STAR standards-based assessments. The STAR results do not, however, include a passing score, so the impact of the HSEE requirements will not be clear.

covers a wide range of proficiency levels from very basic through advanced. One area where clarification would be helpful to districts as well as to state-level policy-makers is whether it would or would not be reasonable to relate the reporting scales for these two programs. For example, can parents and teachers use a student's scores on the STAR exam to anticipate his or her results on the HSEE exam?

In addition to clarifying policy, design, administration, and reporting differences among the different high school testing programs, the Department should consider the feasibility of assembling a database that would allow for comparisons of the scores of individual students on these different examinations. Such a database would provide important empirical information on the relationships among the examinations to complement the policy and design distinctions.

Specific Recommendation 2. The Department and Board should establish, expand, or accelerate processes for communicating with local districts about the HSEE and supporting their preparation for its implementation.

Information from our surveys of principals and teachers and discussions with the panels of educators who participated in our item review workshops indicated a strong desire for information about the HSEE at the district and school level and some confusion about how the new program will be implemented. We believe that it will be important for the state to establish a consistent means of communication with districts and schools that reduces the sense of “insiders” and “outsiders” with regard to the flow of information about the exit examination. By consistent we mean coordination at the state level so different messages from the Department, the State Board, and others do not trickle down to the districts. In designing a communication program, state personnel should seek to anticipate concerns or misunderstandings and provide the district and school leadership with information that can be used to address confusion or turmoil. While there may not always be a definitive answer or resolution that can be offered, an open, straightforward response acknowledges the importance of the educators' stake in the exit examination.

A consistent communication program would also be useful for facilitating communication among the districts, as well as between state and district levels. CDE might work to identify promising practices of “early planners” to provide models for other districts and schools. Examples of district-level practices that might be included are: (a) developing community outreach programs and parent communication systems, (b) implementing effective staff development strategies, and (c) initiating academic supports to students. In addition, information from the surveys indicated a desire for concrete support from CDE with staff development and plans to prepare students, parents, and the community for the initial exam administration.

Specific Recommendation 3. The Department and development contractor need to gather, review, and discuss more information on the appropriateness and effectiveness of testing accommodations for special needs students and English-language learners.

The HSEE Panel has heard presentations on a number of issues relating to testing accommodations. It is likely that the effectiveness of different accommodations in allowing for the inclusion of all students and the appropriateness of the accommodations for generating comparable scores will vary greatly across testing programs. Information on the appropriateness and effectiveness of different accommodations on the HSEE itself is needed. A limited number of special needs students and English-language learners were included in the Spring 2000 field test, but we are not aware of any plans to use field test data to analyze the impact of testing accommodations. If the current schedule is maintained, the fall 2000 field test will be the only opportunity to collect information on testing accommodations prior to the first operational administration. If the schedule is relaxed, there would be more significant opportunities in a spring 2001 test administration.

In collecting information on the effects of different accommodations, it will be important to collect information on the students who are not tested as well as those who are. How do students who cannot be accommodated with available alternatives function in the classroom? What sort of accommodations might be considered in developing different forms of assessment for these students? In addition, it would be highly desirable to include planned variation in the accommodations offered to different students so that the impact of these accommodations on the scores of students without special needs can be investigated. We recognize, however, that this type of study can only be conducted in a research and development mode and not in conjunction with operational testing where administration conditions must be the same for all students.

The appropriateness of testing accommodations for English-language learners is particularly critical, given the significance of this population in California. The use of bilingual dictionaries, for example, may be common in daily instruction for many students and might be a common accommodation in employment or other settings after high school. Nonetheless, the use of such dictionaries may not be judged to be an appropriate accommodation for the HSEE, at least on the Language Arts examination. Clarification of the knowledge and skills to be assessed may be required to explain why this and many other accommodations are or are not allowed.